DOCUMENT RESUME

ED 324 348 TM 015 591

AUTHOR Veccia, Ellen M.; Schroeder, David H.
TITLE Development and Internal-Structure Analysis of a New Pitch Discrimination Test, Worksample 740 A. Technical Report 1990-1.
INSTITUTION Johnson O'Connor Research Foundation, Chicago, IL. Human Engineering Lab.
PUB DATE Jul 90
NOTE 31p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Ability Identification; *Adults; *Aptitude Tests; *Auditory Discrimination; Auditory Tests; Career Counseling; Comparative Testing; Computer Assisted Testing; Construct Validity; Difficulty Level; Factor Analysis; Factor Structure; *Music; Personnel Evaluation; *Talent Identification; *Test Construction; Test Items; Test Reliability
IDENTIFIERS Internal Validity; Music Ability; *Pitch (Music); Pitch Discrimination Test Worksample 740 A

ABSTRACT

As a measure of musical aptitude, a new 90-item Pitch Discrimination Test was developed, and its internal structure was examined. Each of the three sections of the test measures an individual's aptitude for pitch discrimination in a different frequency range using square-wave tones generated by a personal computer. A total of 1,303 examinees, mostly clients taking aptitude tests for career guidance, completed the new worksample 740 A and the current version 315 F. The analyses reported were conducted with a minimum of 1,270 examinees; the exact number of cases for any given analysis varied slightly as a function of missing values on some variables. Representation of both sexes was nearly equal, and the median age of examinees was 26 years. The internal structure of Worksample 740 A was assessed using several methods: (1) comparisons with Worksample 315 F; (2) reliability analysis; (3) factor analysis; (4) item fit analysis; and (5) item difficulty analysis. Obtained reliability coefficients for Worksample 740 A were higher than those of Worksample 315 F. With a few exceptions, all items contributed to the test's reliability. Factor analysis indicated that all sections measured the same underlying construct. Analysis of age and sex indicated that age, but not sex, requires careful consideration in construction of norms for the new test. The new instrument allows the Johnson O'Connor Research Foundation to provide examinees with a more generalizable assessment of their aptitudes for pitch discrimination. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

ROBERT KYLE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED324348

DEVELOPMENT AND INTERNAL-STRUCTURE ANALYSIS OF A NEW PITCH DISCRIMINATION TEST, WORKSAMPLE 740 A

Ellen M. Veccia

and

David H. Schroeder

JOHNSON O'CONNOR RESEARCH FOUNDATION, INC.

Technical Report 1990-1

July 1990

TM015591

COPYRIGHT © 1990 BY JOHNSON O'CONNOR RESEARCH FOUNDATION, INCORPORATED
ALL RIGHTS RESERVED

Development and Internal-Structure Analysis of a New Pitch Discrimination Test, Worksample 740 A

Ellen M. Veccia and David H. Schroeder

ABSTRACT

A new 90-item Pitch Discrimination test, Worksample 740 A, was produced by recording square-wave tones generated by a personal computer. Worksample 740 A is unique in that each of the three sections of the test measures aptitude for pitch discrimination in a different frequency range. This aspect of the new test provides a more generalizable measure of pitch discrimination aptitude. A total of 1,303 examinees completed both the new Worksample 740 A and the current version 315 F.

The internal structure of Worksample 740 A was assessed by several methods of analyses, including comparisons with Worksample 315 F, item difficulty analysis, reliability analysis, factor analysis, and item fit analysis. Worksample 740 A correlated acceptably with 315 F, indicating that both tests measure the same underlying aptitude. Several reliability analyses were conducted both within and between sections of Worksample 740 A. The obtained reliability coefficients were consistently high, with an overall alpha of .904. This reliability well exceeds that of Worksample 315 F.

Item-total correlation values and item fit static values indicated that, with a few exceptions, all items contributed positively to the test's reliability. The factor analysis indicated that the three sections of the test measure the same underlying construct.

Worksample 740 A was reliable across different age groups and for the two sexes. Analysis of sex and age effects, however, indicated that age, but not sex, requires consideration in the construction of norms for Worksample 740 A.

Overall, Worksample 740 A successfully measures pitch discrimination ability in three different frequency ranges and enables the Foundation to provide examinees with a more generalizable assessment of their aptitude for pitch discrimination.

CONTENTS

	Page
Introduction	1
Structure of Worksample 740 A	2
Construction of Worksample 740 A	5
Method	5
Examinees	5
Procedure	6
Results and Discussion	6
Descriptive Statistics	6
Analysis of Worksample 740 A by Subscales	7
Item Analysis of the Full 90-Item Test	10
Reliability Analysis Between Sections	16
Analysis of Sex and Age Effects	16
General Discussion	18
References	20
Appendix	22

ACKNOWLEDGMENTS

We gratefully acknowledge the staffs of the Atlanta, Boston, Chicago, Dallas/Ft. Worth, Denver, Houston, Los Angeles, New Orleans, New York, San Francisco, Tampa, and Washington, D.C., testing offices for their assistance in data collection.

The Research Department staff provided valuable assistance in assuring the accuracy of the data presented in this report.

We thank Richard Gershon and Lore Eargle, who both contributed their efforts to the production of the master tape for Worksample 740 A.

INTRODUCTION

The Pitch Discrimination test is one of three musical aptitude tests in the standard Foundation battery. Each musical test measures ability level for a separate and distinct musical aptitude. Ability to discriminate pitches of different frequencies is an important aptitude for individuals aspiring to careers in music related fields. This aptitude is particularly important if the area of interest involves voice or instruments that do not have fixed pitches, such as many of the stringed instruments (Trembly, 1974).

The earliest version of the Pitch Discrimination test used by the Foundation was the Sense-of-Pitch test developed by Seashore (1919), labeled Worksample 76. Based on data obtained by the Foundation, this early version of the test underwent several revisions by the Foundation staff, resulting in a new version of the test, Worksample 315. This version was then replaced by Seashore's revised version (Saetveit, Lewis, & Seashore, 1940), which was designated Worksample 364 by the Foundation. In 1968, the Foundation began using Worksample 315 again, and it has been the standard worksample for Pitch Discrimination up to the present. The history of the Pitch Discrimination test, as well as a review of its revisions, is available in the test manual for the music tests, *Pitch Discrimination Worksample 315, Rhythm Memory Worksample 366, and Tonal Memory Worksample 498 Manual*.

Worksample 315 has the same basic format as many of the previous forms of the Pitch Discrimination test. Each item of the worksample consists of the presentation of two tones. One of these tones, the reference tone, is 435 Hz. The comparison tones range from 437 to 445 Hz. Examinees are required to indicate whether the second tone of the pair is higher or lower than the first tone. For "higher than" items, the reference tone is presented first, followed by the comparison tone (as indicated above, all comparison tones have a higher frequency than does the reference tone). This order of presentation is reversed for "lower than" items.

One characteristic common to all of the previous Pitch Discrimination tests is that the reference and comparison tones were limited to a single frequency range. For the most part, these tests used reference tones of 435 Hz and comparison tones that ranged from 465 Hz to 435.5 Hz. One consequence of limiting the test to a single frequency range is that interpretation of persons' aptitudes is also limited. This limitation of the current test could be especially problematic for individuals seeking careers in areas of the music industry requiring the ability to discriminate pitch in several different octaves.

To enable the Foundation to provide examinees with a more thorough assessment of their aptitude for pitch discrimination, a new test was constructed. As will be explained, this test measures aptitude for pitch discrimination in three different pitch ranges. It is this characteristic of the test that increases its generality.

The purpose of this report is to present the analysis of the internal structure of the new Pitch Discrimination test, Worksample 740 A. The structure, as well as the actual construction of the test, will be reviewed with comparisons made between this new form and the currently used form, 315 F. Finally, analyses of sex and age effects are presented.

Structure of Worksample 740 A

As discussed by Schroeder (Test Information Bulletin 1989-1), Worksample 740 A differs from the current Worksample 315 F in several ways. Most importantly, Worksample 740 A tests pitch discrimination across three frequency ranges. Worksample 315 F is restricted to a middle frequency range, with the reference tone at 435 Hz and comparison tones ranging from 437 to 445 Hz. Daniel (1982) suggested that a more comprehensive assessment of pitch discrimination ability could be obtained by testing across several pitch ranges. As Daniel suggested, 740 A was constructed using three different frequency ranges. (See also the music manual, *Pitch Discrimination Worksample 315, Rhythm Memory Worksample 366, and Tonal Memory Worksample 498 Manual*, p. 6).

There are three sections in Worksample 740 A, with items numbered progressively from 1 to 90. Each section tests pitch discrimination in a different frequency range. The reference tones for Sections 1, 2, and 3 are 260 Hz, 435 Hz, and 1000 Hz, respectively. Notice that the middle frequency range (435 Hz), is near the note middle C on the piano keyboard, while the lower and higher frequency ranges are about one octave below and above this note, respectively. The 30 items in each section progress from the easiest to the most difficult discriminations. For example, the first item in Section 3 (Item 61) presents a tone pair of 1020 and 1000 Hz. In contrast, the last item of the test, Item 90, presents a tone pair of 1004 and 1000 Hz.

In order to compare the degree of difficulty across the three sections of 740 A, one must take into account the widely accepted finding that the threshold for discrimination between two stimuli is proportional to the value of the reference stimulus. This characteristic of psychophysical discriminations is known as Weber's Law and is expressed mathematically by the following formula:

(1) Formula for Weber's Law $\Delta S = kS$,

where ΔS is the smallest difference between stimulus values that can be discriminated, k is a modality-dependent constant, and S is the value of the reference stimulus.

In terms of the present worksample, Weber's Law specifies that the smallest discriminable difference in pitch is determined by the pitch of the reference tone. For Worksample 740 A, k was determined by dividing the smallest difference in pitch used for Worksample 315 F (2 Hz) by the pitch of the reference tone for this worksample (435 HZ). This resulted in a k = .0046. This constant was then applied to Formula (1) to determine the corresponding discrimination (smallest difference in Hz) for each of the three reference tones. For the low frequency range, ΔS = 1.20 Hz (obtained by multiplying .0046 by 260 Hz), and for the high frequency range ΔS = 4.6 Hz (.0046 multiplied by 1000 Hz). Both ΔS values were rounded down to a whole number. The ranges for the comparison notes for Sections 1, 2, and 3 are as follows: 265-261 Hz for Section 1, 443-437 Hz for Section 2, and 1020-1004 Hz for Section 3. Table 1 presents a summary of the test's content.

As should be clear from Table 1, Section 2 of Worksample 740 A was comparable to Worksample 315 F with one exception. Previous research indicated that frequency differences in excess of 8 Hz did not contribute to the test's ability to detect aptitude differences among individuals in the upper three quartiles of the distribution (Technical Report 86). Therefore, comparison tones of 445 Hz, a 10 Hz difference, were not used in Section 2 of Worksample 740 A. (A more detailed description of the test is shown in the first two columns of Table 2, where the item number and the pitch frequency specifications for each of the 90 items are shown exactly as they are presented on the test.) The three practice items that preceded the test proper were: 260-265 HZ, 265-260 HZ, and 260-265 HZ.

Worksamples 740 A and 315 F also differ with respect to the total number of test items. Worksample 740 A has 90 items, compared to 80 items on 315 F. The reliability for Worksample 315 F was .80 (Statistical Bulletin 1980-15), which barely meets the Foundation's standard for acceptability (Test Information Bulletin 1980-7). It was expected that the additional items on 740 A would result in at least a minimal increase in the reliability of the worksample (Statistical Bulletin 1988-2).

Like previous versions of the Pitch Discrimination test, Worksample 740 A is composed of items with "higher" and "lower" as the correct response. Since all comparison tones are a higher frequency than their respective reference tones, all "higher" items present the reference tone followed by the comparison tone. For "lower" items, the order of presentation is reversed so that the comparison tone occurs first, followed by the reference tone. Worksample 740 A has an equal number of "higher" and "lower"

items. The order of occurrence of "higher" and "lower" items was determined randomly with the restriction that not more than four items of either type occur consecutively (e.g., a maximum of four consecutive items with "higher" as the correct response).

Table 1

Content of Pitch Discrimination Worksample 740 A

Number of items	Reference pitch (Hz)	Comparison pitch (Hz)	Mean % correct	SD (%)
Section 1				
3	260	265	91	21
3	260	263	86	24
12	260	262	82	20
12	260	261	69	19
Section 2				
3	435	443	92	19
3	435	440	88	23
8	435	439	86	19
8	435	438	82	20
8	435	437	78	20
Section 3				
3	1000	1020	88	24
3	1000	1012	85	23
8	1000	1010	85	20
8	1000	1007	85	20
8	1000	1004	74	22

Construction of Worksample 740 A

Production of the Pitch Discrimination worksample has long been problematic and has sometimes required the use of professional recording studios (*Pitch Discrimination Worksample 315, Rhythm Memory Worksample 366, and Tonal Memory Worksample 498 Manual*). Worksample 740 A was produced within the Foundation. Richard Gershon, a research assistant in the Research Department, generated square wave tones by programming a personal computer with a tone generator. As on the previous Pitch Discrimination test, Worksample 315 F, the duration of each tone is .6 seconds with a .5 second interval between the reference tone and comparison tone. There is a 2.0 second pause between items. After a tape of the tones was produced, Lore Eargle, the director of the Tampa testing office, spliced the tones and the instructions together (Test Information Bulletin 1989-1). (The script of the test instructions is provided in the Appendix.)

The use of square waves is unique to Worksample 740 A, but there was no reason to expect that square waves would affect performance. The most notable effect of square waves is the potential for a slight click to occur at the onset of the tone (Thurlow, 1971). None of the pilot subjects indicated that a click occurred, or that there was other interference with the tones, and thus there was no reason to expect any performance effects as a result of the square waves.

METHOD

Examinees

For the 740 A testing, examinees were unselected for sex and age, but scholarship and public information examinees were not included. Examinees' ages ranged from 14 to 64 years, M = 28.48, standard deviation (SD) = 10.36. The median age was 26 years. Sex was nearly equally represented, with males representing 51.6% of the sample and females 48.4%. Geographically, about 40% of the examinees were tested in the Foundation's eastern offices (Boston, New York, and Washington, D.C.), about 34% in the southern and southwestern offices (Atlanta, Dallas/Ft. Worth, Houston, New Orleans, and Tampa), and approximately 26% were tested in the Foundation's western and midwestern offices (Chicago, Denver, Los Angeles, and San Francisco). Examinees were generally white and characterized economically as middle to upper-middle class. The examinees typically tested by the Foundation have at least some college education or are college-bound. Examinees were paying clients of the Foundation and, as

such, self-selected for aptitude testing, usually for the purpose of career guidance.

Procedure

Worksample 740 A was administered from February through April 1989, in all testing offices except Seattle. Because this test was administered on an experimental basis, examinees completed both Worksamples 315 F and 740 A. This new Pitch Discrimination test was administered only after examinees had completed their first appointment (when they took 315 F).

The administration procedures for 740 A are similar to those of the current Pitch Discrimination test. Worksample 740 A is administered by tape through headphones. Testing offices were provided with the experimental tape, as well as the 90-item answer sheets. As with the previous test, for each tone pair, examinees indicated whether the second tone was higher or lower than the first tone (see Appendix A).

Answer sheets for Worksample 740 A were returned to the Research Department for scoring and analysis. Thus, examinees were not provided with a score for this test. Unless indicated otherwise, the analyses reported here were conducted with a minimum of 1,270 examinees. The exact number of cases for any given analysis varied slightly as a function of missing values on some variables.

RESULTS AND DISCUSSION

The test is scored by assigning a score of 1 to items that are answered correctly and a score of 0 to items answered incorrectly (or omitted). Total test score is the sum of the item scores.

Descriptive Statistics

The mean raw score on the 90-item test was 73.29, SD = 13.48. Scores ranged from 36 items correct to 90 items correct with a median score of 77 items correct. In terms of percent correct, the mean percent correct on the 90-item test was 81.44 with an SD = 14.98. Scores ranged from 40% to 100% correct with the median at 85.5. In comparison, the mean percent correct on Worksample 315 F was 78.58 with an SD of 12.24. Analysis of the difference between mean percent correct on 740 A and 315 F yielded a $t(1277) = 10.02$, $p < .001$; with an effect size $d = .21$. The d obtained here is considered a small effect size (Cohen, 1988). Thus, the new Pitch Discrimination test is a little

easier than the old test. Since the effect size for the difference in scores on Worksamples 740 A and 315 F was small, however, there is no reason to expect that there are important differences in the nature of the aptitude measured by these worksamples.

The mean (*M*) percent correct for Section 1 (low frequency range) was 77.82, *SD* = 16.35. The mean for Section 2 (middle frequency range) was 83.72, *SD* = 16.13, and for Section 3 (high frequency range), 82.58, *SD* = 16.81. Inspection of these means suggested that Section 1 was more difficult than the other two sections. This was confirmed by performing a repeated measures analysis of variance (ANOVA) on section scores and calculating the estimated effect size (ω^2). The results of this analysis yielded $F(2, 2600) = 184.41$, $p < .001$. The main effect of section accounted for about 12% of the total variance ($\omega^2 = .124$), an effect size that is comparable to a correlation of approximately .35 (ω). This value of omega represents a medium-to-large effect (Cohen, 1988), and requires further comment. It seems, however, more appropriate to undertake this comment after the remainder of the analyses have been presented. Therefore, we will address this finding in more detail in the general discussion.

Since Section 2 is comparable to Worksample 315 F, this section provides the most direct comparison between the earlier form and the current worksample. The mean percent correct for Section 2 of Worksample 740 A was approximately 5.2% higher than the mean percent correct for Worksample 315 F. This difference is significant, $t(1289) = 15.68$, $p < .001$, and yields a small-to-medium effect size, $d = .36$. It is interesting to note that Section 2 should actually have been more difficult than Worksample 315 F because the easiest discriminations (435-445 Hz) were not included in 740 A. As above, this unexpected finding will be addressed in more detail in the general discussion.

Analysis of Worksample 740 A by Subscales

Worksample 740 A is the first Pitch Discrimination test containing reference and comparison pitches across several ranges of frequencies. In order to fully understand this characteristic of the test, extensive reliability analyses were conducted in addition to the more standard analyses that will be presented later in this report. All reliability coefficients presented here are Cronbach's alpha unless otherwise stated.

The alpha reliability coefficient is influenced by the strength of the relationships between items on a test. In other words, the correlations between test items are an important determinant of reliability coefficients. Because of this relationship between inter-item correlations and the reliability

coefficient, one must exercise caution in the conduct of reliability analyses on tests containing repeated items. That is, items that are identical to each other are likely more strongly correlated than items that are similar, but not identical. The magnitude of the correlation between these identical items may falsely inflate the reliability coefficient.

Worksample 740 A contains a number of identical items for each of the specific tone pairs. As discussed above, the repetition of identical items within a single test may falsely inflate reliability measures. In order to determine the most accurate reliability coefficient for 740 A, an analysis that eliminates this potential complication was employed. Since the occurrence of repeated items is restricted to specific reference-comparison tone pairs, a single score was calculated for all items for a specific tone pair. This procedure resulted in a separate score for each of 14 subscales for Worksample 740 A, with each subscale score based on items that were contained only in that particular subscale. The last two columns of Table 1 show the mean percent correct and *SD* for each of the subscales. These 14 scores were subjected to a reliability analysis for a 14-unit test with each subscale score representing one unit.

The alpha coefficient for the 14 subscales was .904 with a mean inter-subscale correlation of .472. The standard error of measurement (*SEM*) was 4.18 raw score points for the 90-item test. This is a high value for a reliability coefficient and indicates that scores across the subscales of Worksample 740 A are consistent, which, in turn, indicates high internal consistency for the test. In addition, this alpha coefficient substantially exceeds both the Foundation's minimum level of acceptability (.80) and the alpha coefficient obtained for Worksample 315 F (.80).

The Spearman-Brown formula may be used to calculate the expected reliability for tests of comparable content but different numbers of items. Since Worksamples 740 A and 315 F should be roughly comparable, the Spearman-Brown formula may be appropriately applied here. The reliability coefficient for the subscales of Worksample 740 A represents a substantial increase over the .80 alpha coefficient reported for the 80 items of Worksample 315 F (Statistical Bulletin 1980-15). By the Spearman-Brown formula, the increase in the number of items (from 80 to 90) should have increased the obtained reliability from .80 to .82. This was obviously not the case, as the obtained reliability was a great deal higher than .82.

This unexpectedly large increase in reliability occurred both for this conservative analysis by subscales, and for the more liberal analyses which will be presented later in this report. This unexpected increase will be addressed further in the general discussion. At this point, however, a comment on the

consistency of the magnitude of alpha is in order. The consistently large magnitude of alpha obtained for Worksample 740 A indicates that this worksample has characteristics, other than its number of items, that contribute strongly to the reliability of the test.

Analysis of subscales within sections. The reliability of Worksample 740 A was further examined by determining the alpha coefficient for the subscales *within* each section. Once again, for these analyses, subscales were used as the units of analysis with Section 1 having 4 units, and Sections 2 and 3 having 5 units each. The alpha coefficients for Sections 1, 2, and 3 were .683, .801, and .802, respectively. For each section, the Spearman-Brown formula was used to determine the expected reliability for a 90-item test composed entirely of items within that section. Increasing the number of items in Section 1 from 30 to 90 would be expected to yield a reliability coefficient of .866. Comparable calculations for Sections 2 and 3 resulted in expected reliability coefficients of .923 (Section 2) and .924 (Section 3).

The expected reliability coefficients for Sections 2 and 3 are substantial, indicating that items with reference tones of 435 Hz and 1000 Hz are very reliable. Compared to Sections 2 and 3, the expected reliability coefficient for Section 1 is somewhat lower, though still falling above the Foundation's minimum acceptable level (.80). The results for the reliability analyses within sections are consistent with the findings already presented in that Section 1 seemed to behave somewhat differently from the other two sections of Worksample 740 A.

Analysis of subscales by sex and age. This section of the report presents reliability analyses (Cronbach's alpha) conducted on subgroups of the sample using the 14 subscale scores as units. These analyses served as verification that the reliability of Worksample 740 A was consistent both within and between sex and within and between different age groups.

The first reliability analysis on subgroups of the data examined the reliability of 740 A within each sex. The alpha coefficient for the 14 subscales across female examinees (N = 621) was .904, and the alpha across male examinees (N = 663) was .903. The magnitude of these alpha coefficients indicates that 740 A is equally and substantially reliable for both females and males.

Next, examinees were grouped into nine different age groups, with females and males combined, and reliability was assessed across these groups. The age groups were identical to those used for the analysis of sex and age effects (see below). The sample sizes for age groups ranged from 80 to 168. The alpha coefficients for the nine age groups ranged from .891 to .914

with a mean of .902. The magnitude of the alpha coefficients that resulted from these analyses indicates that the reliability for Worksample 740 A is about equal across different age groups.

Finally, examinees were grouped into the same nine age groups used above, but the age groups were analyzed separately for each sex. For females, the sample sizes of the age groups ranged from 37 to 91. The reliability coefficients computed for females in each age group ranged from .874 to .916 with a mean of .899. For males, the sample sizes of the age groups ranged from 43 to 91. The reliability coefficients computed for males in each age group ranged from .877 to .911 with a mean of .901. Once again, the coefficients obtained here indicate that the reliability of Worksample 740 A is comparable across females of differing ages and males of differing ages.

To summarize, the results reported here for the reliability of Worksample 740 A by sex and age are consistent with the 14-subscale alpha coefficient reported earlier. In addition, these findings indicate that Worksample 740 A is a reliable measure of aptitude for pitch discrimination for both sexes and for all the age groups typically tested by the Foundation.

Item Analyses of the Full 90-Item Test

This section of the report presents the reliability analyses for Worksample 740 A across the 90 items of the test. This series of analyses was modeled after those reported by Bethscheider (Technical Report 1988-4) and included calculation of Cronbach's alpha, an index of item difficulty, item-total correlations, Rasch-model infit measures (Wright, Linacre, & Schultz, 1989), and factor analysis. In the Rasch model, the infit statistic is an index of item quality in terms of how well the item discriminates ability level. An average item has an infit value between +2 and -2. Exceptionally good items have large negative infit values, while extremely poor items have large positive infit values. As discussed earlier, because identical items occur on each subscale, a reliability coefficient that treats each item as independent is likely to be somewhat inflated. The more appropriate reliability analysis was calculated using subscale scores as the units of analysis (see previous section on subscale reliability).

Item analysis of the full 90 items. After the reliability of subscale scores was examined, the next step in assessing the reliability of Worksample 740 A required item-level analyses of the 90-item test. Table 2 displays a detailed listing of the 90 items, including frequency specifications, difficulty values, corrected item-total correlations, and values of the Rasch infit statistic. Cronbach's alpha for Worksample 740 A (90 items) was .939 with a mean inter-item correlation of .156. *Again, this is an inflated estimate of the test reliability because of the*

Table 2

Item Analysis for Pitch Discrimination Worksample 740 A

Item	Item frequency specification (Hz)	Difficulty (*p* value)	Item-total[a] correlation	Fit statistic value
1	265, 260	.91	.40	-1.19
2	265, 260	.91	.40	-1.17
3	260, 265	.91	.41	-1.37
4	263, 260	.92	.39	-1.10
5	260, 263	.84	.43	-1.50
6	260, 263	.83	.41	- .78
7	260, 262	.69	.40	1.53
8	262, 260	.88	.42	-1.48
9	262, 260	.87	.42	-1.35
10	260, 262	.78	.45	-1.44
11	262, 260	.86	.42	-1.34
12	260, 262	.78	.39	.19
13	260, 262	.73	.41	.55
14	262, 260	.88	.41	-1.24
15	262, 260	.84	.41	- .98
16	260, 262	.77	.44	-1.10
17	262, 260	.87	.40	- .95
18	262, 260	.84	.38	- .15
19	260, 261	.68	.34	4.10
20	260, 261	.61	.19	10.95
21	261, 260	.72	.31	4.03
22	260, 261	.68	.23	7.96
23	261, 260	.78	.36	1.27
24	261, 260	.74	.34	2.67
25	260, 261	.67	.25	7.63
26	261, 260	.75	.29	3.86
27	261, 260	.70	.26	6.43
28	260, 261	.65	.25	7.54
29	260, 261	.55	.15	14.01
30	261, 260	.71	.22	7.42
31	443, 435	.89	.50	-2.60
32	435, 443	.94	.39	-1.11
33	443, 435	.95	.38	- .97
34	440, 435	.83	.48	-2.52
35	435, 440	.89	.42	-1.41
36	440, 435	.90	.45	-1.98
37	439, 435	.83	.47	-2.32
38	435, 439	.88	.38	- .86
39	439, 435	.87	.43	-1.67
40	439, 435	.83	.47	-2.18
41	435, 439	.88	.40	-1.14
42	439, 435	.88	.44	-1.78
43	435, 439	.88	.42	-1.29
44	435, 439	.84	.41	- .99
45	435, 438	.77	.36	1.21

Note. $N \geq 1,270$.

[a]All item-total correlations significant beyond $\alpha = .01$.

Table continues

Table 2 (cont'd)

Item Analysis for Pitch Discrimination Worksample 740 A

Item	Item frequency specification (Hz)	Difficulty (*p* value)	Item-total* correlation	Fit statistic value
46	438, 435	.86	.40	- .97
47	435, 438	.80	.39	- .22
48	438, 435	.84	.47	-2.23
49	435, 438	.83	.39	- .48
50	438, 435	.85	.40	- .90
51	435, 438	.83	.44	-1.63
52	435, 438	.80	.35	.87
53	435, 437	.72	.23	6.45
54	437, 435	.83	.43	-1.26
55	435, 437	.75	.30	3.74
56	437, 435	.81	.41	- .69
57	435, 437	.69	.28	5.77
58	435, 437	.70	.23	7.41
59	437, 435	.95	.34	- .67
60	435, 437	.81	.34	1.19
61	1020, 1000	.88	.51	-2.80
62	1020, 1000	.89	.45	-1.88
63	1020, 1000	.87	.46	-2.12
64	1012, 1000	.72	.47	-1.23
65	1000, 1012	.93	.35	- .76
66	1000, 1012	.91	.38	-1.05
67	1000, 1010	.88	.38	- .84
68	1010, 1000	.85	.44	-1.77
69	1000, 1010	.87	.40	- .97
70	1010, 1000	.85	.50	-2.84
71	1010, 1000	.85	.47	-2.11
72	1010, 1000	.82	.48	-2.55
73	1010, 1000	.81	.48	-2.34
74	1000, 1010	.91	.37	- .95
75	1000, 1007	.85	.39	- .73
76	1007, 1000	.80	.47	-2.22
77	1000, 1007	.87	.42	-1.44
78	1000, 1007	.86	.34	.11
79	1007, 1000	.83	.50	-2.93
80	1000, 1007	.87	.39	- .79
81	1000, 1007	.88	.37	- .63
82	1000, 1007	.87	.35	- .38
83	1004, 1000	.75	.41	.48
84	1000, 1004	.78	.28	3.38
85	1004, 1000	.73	.39	1.25
86	1000, 1004	.77	.27	3.74
87	1000, 1004	.73	.17	8.13
88	1004, 1000	.72	.41	.81
89	1000, 1004	.74	.26	5.23
90	1004, 1000	.70	.38	2.34

Note. $N \geq 1,270$.

*All item-total correlations significant beyond $\alpha = .01$.

repeated items. The magnitude of this reliability coefficient confirms the expectation that the 90 items on the test all measured the same attribute.

Item difficulty. Item difficulty was determined by *p*. This statistic indicates the proportion of individuals who responded correctly to a given item. The mean *p* value across the 90 items was .81. This value corresponds to the relatively high mean percent correct obtained for the test. Note that the interpretation of both the average *p* value and the mean percent correct should be guided by the fact that since the items were dichotomous, every item should have a minimum *p* value of .50 by chance alone. Similarly, individuals could have obtained a mean total score of 50% (raw score = 45) by chance.

It is not uncommon for examinees to demonstrate a response bias when they are required to respond to items by choosing an answer from recurring options. Here, examinees were required to choose "higher than" or "lower than" for each of the 90 test items. There are 45 items with "higher" as the correct choice, and 45 with "lower" as the correct choice. If response bias was not a factor, then both the *p* values, and the mean percent correct should be about equal for "higher" and "lower" items.

The *p* value for "higher" items ranged from .55 to .94 with the median falling at .81. For "lower" items, the median *p* value was .85 with a range from .70 to .95. The difference in median *p* values for "lower" and "higher" items suggested that item difficulty values were influenced, in part, by whether "higher" or "lower" was the correct choice. This possibility was examined by comparing the mean percent correct for "higher" and "lower" items, although this analysis is only appropriate after the items have been equated for difficulty. Since difficulty varies as a function of subscale, this was accomplished by selecting an equal number of "higher" and "lower" items from each subscale of the test. This selection procedure resulted in 35 "higher" items and 35 "lower" that were included in the analysis. "Higher" items had a mean percent correct of 78.8 (SD = 16.5), and "lower" items had a mean percent correct of 81.2 (SD = 18.5). This difference was significant, $t(1287) = 6.08$, $p < .001$, and yielded a medium-to-small effect size (d = .41; Cohen, 1988).

Since there are equal numbers of "higher" and "lower" items on the test, and since "high" and "low" items were equated for difficulty, the results of the analysis presented above indicate that examinees had a response bias for "lower" as the correct response. In other words, other things being equal, examinees were more likely to select "lower" as the correct response. In terms of the reliability of the test, this response bias is not problematic, but it may have influenced the results of the factor analysis that will be presented later in this report.

Item-total correlations. Corrected item-total correlations indicate how well items correlate with total scores on the test. The correction feature removes the item under consideration from the total score in order to prevent inappropriate inflation of the item-total value. Typically, items with the highest item-total correlations contribute the most positively to reliability, while those with low item-total correlations can reduce reliability (Ghiselli, 1964). Thus, there is a relationship between item-total correlation values and reliability coefficients.

The corrected item-total correlations for Worksample 740 A ranged from .15 to .51 with a mean of .38. The item-totals shown in Table 2 indicate that most items contributed positively to the test's reliability by demonstrating acceptable (above .30) corrected item-total correlations. There was an interesting relationship between item-total correlation and item difficulty (*p*). Items that were more difficult tended to have lower item-totals than did easier items. Essentially, these low item-total correlations indicate that some individuals responded correctly to items that were more difficult than their overall ability level, as indicated by their total score. It seems reasonable to attribute this discrepancy to the chance factor discussed earlier. That is, individuals with low pitch discrimination aptitude had a 50% chance of responding correctly to difficult items simply by guessing.

Fit statistic values. Rasch fit statistics were used as an additional index of item quality. The infit value for each item is shown in the last column of Table 2. Most of the infit values fall in the acceptable range, indicating that the Worksample 740 A items adequately discriminated different ability levels. Items with the poorest (high positive) infit values tend to be the most difficult items (e.g., Item 29). As with the item-total correlations, the infit values for difficult items likely reflect the chance factor. That is, some individuals with low ability *guessed* the correct answer to items that were beyond their ability level. Overall, items with low item-total correlations also tended to have high positive infit values (see Table 2).

The majority of the items on Worksample 740 A have acceptable *p* values, item-total correlations, and fit statistic values. Still, these indices are especially poor for some of the items. For example, Items 20, 22, 25, 28, 29, 30, 58, and 87, have high positive infit values, low item-total correlations, and relatively low *p* values. From a statistical perspective, these items are "worse" than the other items on the test, but one must exercise caution in reaching such a conclusion.

Recall that the item specifications are identical for groups of items (the 14 subscales referred to earlier). Thus, for each "poor" item listed above, there are a number of *identical* items

that are "good" items. To ensure that these items had the same subjective quality as other test items, the master and several copies of Worksample 740 A tapes were carefully screened. There was no detectable difference (with the exception of Item 29) between these items and the other items on the test. Since the "poor" items are essentially the same as the "good" items with the same specifications, the variation in the indices used here probably reflects random factors, rather than variation in the actual quality of the items. Item 29 is an exception in that the subjectively judged quality of this item is different from the other items on the test because of an irregularity in the first tone.

Factor analyses. An exploratory principal components factor analysis was conducted on the 90 items in order to determine the degree to which the items measured the same underlying construct (presumably aptitude for pitch discrimination). This analysis resulted in 24 factors with eigenvalues greater than 1.0; there were only two factors, however, that accounted for a percent of variance exceeding 2%. Factor 1 accounted for 17.4% of the total variance, and Factor 2 accounted for 4.6% of the total variance.

The results of this analysis were examined further by employing an objective analysis of the scree plot, as well as a more subjective visual inspection of the same. Discontinuity analysis yields an objective indication of where the scree begins by calculating the slope of the line connecting each of the eigenvalues (Coovert & McNelis, 1988). The scree plot itself is simply a graphical representation of the eigenvalues. Scree refers to the point at which the eigenvalues begin to cluster--in other words, the point where there appears to be little change in the slope of the line between successive eigenvalues.

The discontinuity analysis indicated that slope changes beyond the third eigenvalue were negligible. Visual inspection of the scree plot confirmed the discontinuity analysis. That is, the scree began at the third eigenvalue, indicating that a two-factor solution was appropriate.

To aid the interpretation of a two factor solution, a second principal components analysis was conducted that specified the extraction and varimax rotation of two factors. Inspection of the unrotated factor loadings indicated that Factor 1 represented a general factor with all items having fairly high loadings on this factor. Thus, the unrotated Factor 1 likely reflects the general underlying construct of aptitude for pitch discrimination. For the unrotated Factor 2, items with "higher" as the correct response had positive factor loadings, while items with "lower" as the correct response had negative factor loadings. Thus, the unrotated Factor 2 represents the degree to which persons are biased toward responding "higher" (resulting in high Factor 2 scores) or "lower" (leading to low Factor 2 scores).

When the two factors were varimax rotated, the general factor, as usually happens with varimax rotation, was split into two factors: performance on the "lower" items and performance on the "higher" items. This suggests that, to some extent, the persons who were biased on the "lower" items were distinct from the persons who were biased on the "higher" items.

In general, however, as indicated by the unrotated solution, the factor analysis indicates that Worksample 740 A contains a reasonably strong general factor, representing general pitch discrimination ability, and a smaller additional factor representing persons' differing biases toward responding with "higher" or "lower." As noted earlier, the majority of examinees seemed to have a mild bias toward responding "lower."

The factor analysis also indicated that items in the three pitch ranges all load on the same factor, implying that they measure the same underlying construct.

Reliability Analysis Between Sections

Although the finding of two factors in the factor analysis was reasonably attributed to response bias, to further ensure that the three sections of Worksample 740 A all measure the same construct, an additional reliability analysis was conducted. If, as expected, the three sections of 740 A measure the same construct, then an acceptable reliability coefficient calculated by comparing the scores *between* the three sections of the test should obtain. This analysis was conducted by using scores on Sections 1, 2, and 3 rather than scores on individual items. This effectively resulted in computing the reliability for a three-unit test with each unit consisting of a section score. The alpha coefficient for the three section scores was .901 with a mean inter-item correlation of .753. The magnitude of this coefficient confirms the expectation that even though the three sections of 740 A differ with respect to their frequency ranges, they were consistent with respect to the aptitude measured. This serves as an additional confirmation of the test's reliability and a confirmation for the interpretation of the factor analysis provided above.

Analysis of Sex and Age Effects

Potential differences on Worksample 740 A total score (90 items) as a function of sex and age were examined by a two-way ANOVA for unequal cell sizes. This analysis of sex and age effects was conducted with the regression approach to ANOVA. The regression approach differs from the classical experimental approach by assessing the effect of each variable independent of the influence of all other variables (including interactions) present in the analysis. That is, the effect of age is assessed after the effect of sex and the sex by age interaction are removed, and the effect of sex is assessed after the effect of

age and the sex by age interaction are removed. Thus, this is a more conservative analysis of sex and age effects than is the classical experimental approach.

The nine age groups for females and males used in this analysis were: (a) 14-16, (b) 17-18, (c) 19-20, (d) 21-23, (e) 24-26, (f) 27-30, (g) 31-35, (h) 36-41, and (i) 42 and older. Omega-squared (ω^2), an index of the proportion of variance accounted for by a given effect, was calculated for the main effects of sex and age, as well as the sex by age interaction.

The ANOVA described above yielded a significant main effect for age, $F(8, 1265) = 6.10$, $p < .001$. For the age effect, $\omega^2 = .04$, indicating that age accounted for 4% of the variance in scores. The ω^2 obtained here is comparable to a correlation of .19 between age and score and is considered a small-to-medium effect size (Cohen, 1988).

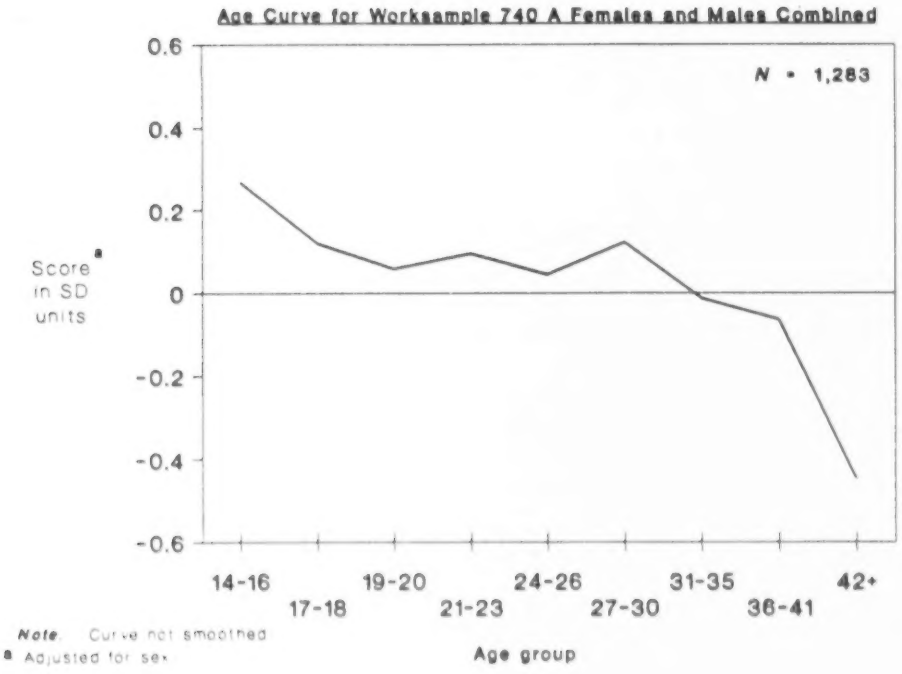
The age curve is shown in Figure 1. In this figure, scores in *SD* deviation units for each of the nine age groups are adjusted for the effect of sex. Figure 1 shows that scores were fairly stable for examinees 17 to 41 years of age. Compared to these age groups, examinees in the youngest age group (14-16) had higher scores.

Since examines in the youngest age group performed as well (or better) than older examinees, it appears that aptitude for pitch discrimination is mature by the mid-teens. That is, there is no increase in aptitude for pitch discrimination as a function of increasing age. The stability in scores for examinees from 17 to 41 years of age indicates that the adult plateau continues through 41 years of age.

Scores for examinees in the oldest age group (42+) showed a sharp decline compared to the younger age groups, but additional data are required before the decline in performance as a function of age can be estimated accurately.

The main effect of sex was significant, $F(1, 1265) = 13.18$, $p < .001$, with males ($M = 74.72$) scoring higher than females ($M = 71.75$). Omega-squared for the sex effect was .01, indicating that sex accounted for only 1% of the variance in scores. This is comparable to a correlation of .10 between sex and score. The size of the sex effect was in the small range (Cohen, 1988) with $d = .22$, indicating that the male mean was about one-quarter of a standard deviation higher than the female mean. The sex by age interaction was also significant $F(8, 1265) = 2.45$, $p < .05$. For the interaction of sex by age, $\omega^2 = .01$, indicating that the interaction also accounted for only 1% of the variance in scores. Both the main effect of sex and the sex by age interaction yielded statistically significant results, but neither accounted for a sufficient proportion of variance in scores to warrant further consideration.

Figure 1



GENERAL DISCUSSION

Several of the analyses reported here indicate that Section 1 is somewhat different from the other two sections of 740 A. One of the ways in which Section 1 appears to differ is in terms of difficulty. The reasons for this difference are not clear, but it may be a function of at least two factors. First, the practice items for 740 A may have been too difficult, thereby affecting performance on the section that immediately follows these items, Section 1. The item specifications for the practice items were identical to those for the first portion of Section 1. The item analysis for Section 1 indicated that the early items in this section were easy for most, but not all, people. If the difficulty of the practice items did affect performance on Section 1 for some examinees, this problem can be easily alleviated. When new tapes are produced for this worksample, the difference in frequency between the comparison and reference tones of the practice items should be increased from 5 Hz to 10 Hz.

Second, in terms of detection threshold, low frequency tones require more volume for detection than do higher frequency

tones (Thurlow, 1971). The tones in Section 1 of Worksample 740 A fall within this low frequency range. Thus, when a new tape is produced, care should be taken to compensate for this relationship. That is, the tones in Section 1 should be produced at a decibel level such that the perceived loudness of these tones is equal to that of the tones in Sections 2 and 3. Note, however, that in spite of the differences in performance on Section 1 versus Sections 2 and 3, the expected reliability for a 90-item test composed only of Section 1 items and the reliability for the three sections of the test combined exceeded the minimum acceptable level.

Comparison between Worksamples 740 A and 315 F indicated that the two forms correlated well with each other. Still, performance on Worksample 740 A was significantly higher than that on Worksample 315 F. This difference is surprising considering that 740 A actually has a larger proportion of more difficult discriminations than does Worksample 315. Once again, the reason for this difference is not clear. The fact that Worksample 740 A has three different sections may reduce the monotony of this type of test, and thereby enhance examinees' attention. In addition, several of the pilot examinees reported Worksample 740 A to be a more pleasant test than Worksample 315.

To summarize the results presented here, the reliability for Worksample 740 A was assessed using several different analyses. These analyses all resulted in high alpha coefficients. The most accurate reliability estimate was .904. Further, the reliability of Worksample 740 A well exceeds the reliability of the currently used Worksample 315. These reliability analyses, together with the other analyses reported here, indicate that the three sections of the test are all acceptably reliable. Thus, Worksample 740 A reliably measures pitch discrimination across three different frequency ranges.

In terms of face validity, Worksample 740 A measures ability for pitch discrimination in three different frequency ranges. This characteristic of the test enables the Foundation to provide examinees with a more generalizable assessment of their pitch discrimination ability. Of course, from an empirical point of view, validation of Worksample 740 A requires further study.

REFERENCES

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Coovert, M. D., & McNelis, K. (1988). Determining the number of common factors in factor analysis: A review and program. *Educational and Psychological Measurement, 48*, 687-692.

Daniel, M. (1982). Unpublished document, Johnson O'Connor Research Foundation, Fort Worth.

Ghiselli, E. E. (1964). *Theory of psychological measurement*. New York: McGraw-Hill.

Minutes of the Research Planning Meeting. (1990, May). D. Schroeder. Unpublished document, Johnson O'Connor Research Foundation, Chicago.

Pitch Discrimination Worksample 315, Rhythm Memory Worksample 366, and Tonal Memory Worksample 498 manual. (1988). A. M. Sabates & K. H. Kaplan. Chicago: Johnson O'Connor Research Foundation.

Saetveit, J., Lewis, D., & Seashore, C. (1940). Revision of the Seashore measures of musical talent. *Series on aims and progress in research, No. 388*. Iowa City, IA: University of Iowa Press.

Seashore, C. E. (1919). *The psychology of musical talent*. New York: Silver, Burdett and Co.

Statistical Bulletin 1980-15. *Reliability of Pitch Discrimination, Wks. 315 FA*. M. Daniel. Boston: Johnson O'Connor Research Foundation.

Statistical Bulletin 1988-2. *JOCRF test reliabilities and interpretation of test scores*. D. Schroeder. Chicago: Johnson O'Connor Research Foundation.

Technical Report 86. (1940). *An analysis leading to a proposed revision of the Pitch-Discrimination test, Worksample 76*. C. R. Wilks. New York: Johnson O'Connor Research Foundation.

Technical Report 1988-4. *The perceptual abilities project*. J. K. Bethscheider. Chicago: Johnson O'Connor Research Foundation.

Test Information Bulletin 1980-7. *Reliabilities and standard errors of laboratory worksamples*. M. Daniel. Boston: Johnson O'Connor Research Foundation.

Test Information Bulletin 1989-1. *Collecting norm data for the new Pitch test, Wks. 740 A**. D. Schroeder. Chicago: Johnson O'Connor Research Foundation.

Thurlow, W. R. (1971). Audition. In J. W. Kling & L. A. Riggs (Eds.), *Experimental psychology* (pp. 223-271). New York: Holt, Rinehart, Winston.

Trembly, D. (1974). *Learning to use your aptitudes*. San Luis Obispo, CA: Erin Hills.

Wright, B. D., Linacre, J. M., & Schultz, M. (1989). *A user's guide to Bigscale: Rasch-model rating scale analysis computer program*. Chicago: MESA.

Appendix

Instructions for Pitch Discrimination Wks. 740 A

Following is the text for the taped instructions for Worksample 740 A:

Next is Pitch Discrimination, trying to hear the difference between two sounds close to one another in pitch. Take the sheet that says Pitch Discrimination, Worksample 740, in the upper left-hand corner, and write your name in the space provided in the upper right-hand corner.

You will hear a musical tone and then another musical tone just after it. Listen to them carefully, and decide whether the second tone is higher than the first or lower than the first. Do not write anything. Just listen. -----

The second tone was higher than the first. Now listen to another pair of tones. Again decide whether the second tone is higher than the first or lower than the first. ----- That time the second tone was lower than the first. Try one more pair for practice, and once again decide whether the second tone is higher than the first or lower than the first. ----- That time the second tone was higher than the first.

Now look at your answer sheet. At the top are three boxes labeled "Practice" for you to mark the answers to three practice pairs of tones. The first box at the left is for your answer to the first practice pair; the second box, in the middle, for your answer to the second pair; and the third box, at the right, for your answer to the third pair.

Notice that each box is divided in half, with an "H" in the upper half, and an "L" in the lower half. Circle the "H", which stands for higher, if the second note is higher than the first; circle the "L", which stands for lower, if the second note is lower than the first. Press firmly with your pencil, and draw a neat circle. Here are the three practice pairs. -----, -----, -----.

Your answers should be H-L-H. In the first practice pair, the second note was higher than the first, and you should have circled the letter "H" in the first of the three boxes. In the second practice pair, the second note was lower than the first, and you should have circled the letter "L" in the second practice box. In the third practice pair, the second note was higher than the first, and you should have circled the letter "H" in the third practice box.

This worksample consists of a number of pairs of notes like

those you have just heard. In each of the pairs, the second note will be either higher or lower than the first. And for each pair, you are to show whether the second note is higher than the first or lower than the first by circling either "H" for higher or "L" for lower in the box for that pair of notes. Use the boxes from left to right across each row.

As the worksample goes on, the difference between the notes becomes smaller and smaller, until you may not be able to hear any difference at all. But there is always a difference. Please guess if you are not certain, even if you have no idea. In this way, you will always be able to keep your place.

If you wish to change an answer, erase your first answer and then circle your new answer.

Although you may change your answers, do not puzzle too long over any one. There are many pairs of notes in this worksample, and you must be able to give your full attention to each pair as you hear it. Your first reaction to the notes, when they are freshest in your mind, is probably your best guide.

Remember: drawing firmly and neatly, circle "H" if the second note is higher than the first, circle "L" if the second note is lower than the first. Concentrate on each pair of notes as you hear it; do not puzzle too long over any one pair.

Get ready for the first row, Row 1. -----
Here is Row 2. -----
Now begin Row 3. -----
Here is Row 4. -----

As you will notice, the notes in the next section of this worksample are in a different range than those in the section you have just completed. Again, your task is to decide whether the second note in each pair is higher or lower than the first and to mark the corresponding answer. The differences between the notes in this section again will be larger at first and then become smaller and smaller. Be sure to mark an answer for each pair of notes, even if you cannot hear any difference between the notes.

Now, begin Row 5. -----
Here is Row 6. -----
Now start Row 7. -----

The notes in the next section are in yet another range than those in the first or second sections of the worksample. Again the differences between the notes will change from larger to smaller differences, but you should be sure to mark an answer for each pair of notes.

Now, begin Row 8. -----

Here is Row 9. -----
And the last row, Row 10. -----

This is the end of this worksample. Put the answer sheet to one side.

Pitch Discrimination is the ability to distinguish fine differences in pitch. It is most necessary in playing instruments for which the performer has to find the right pitch for a note: for the trombone and violin, for example, and for singing. It is somewhat less crucial for playing fixed-pitch instruments like the piano, for which the performer need only strike the right key to get the right pitch.

JOHNSON O'CONNOR RESEARCH FOUNDATION
HUMAN ENGINEERING LABORATORY